Indian J. Genet., 67(4): 329-333 (2007)

Development of highly polymorphic SSR markers for chickpea (*Cicer arietinum* L.) and their use in parental polymorphism

S. A. Qadir, S. Datta, N. P. Singh and Shiv Kumar

Division of Crop Improvement, Indian Institute of Pulses Research, Kanpur 208 024

(Received: August 2007; Revised: November 2007; Accepted: November 2007)

Abstract

Simple sequence repeat (SSR) or microsatellite marker is currently the most preferred molecular marker system owing to their highly desirable properties *viz.*, abundance, hyper-variability, and suitability for high-throughput analysis. Development of SSR markers using molecular methods is time consuming, laborious, and expensive. Use of computational approaches to mine ever-increasing sequences such as expressed sequence tags (ESTs) and genomic DNA sequences available in public databases permits rapid and economical discovery of SSRs. Because the number of SSR markers currently available in chickpea is very limited, the aim of this study was to develop and characterize more SSR markers. Twenty one hundred DNA sequences of chickpea were searched for SSRs and analyzed for the design of PCR primers amplifying the SSR reach regions. Di-nucleotide repeats were found to be the most abundant followed by tri- or mono-nucleotide repeats. The motifs A/T, GA/AG/CT/AC/TC/CA/TA, and CAA/TCT/AGA/CAG/TTG/ATT were the predominant mono-, di-, and tri-nucleotide SSRs, respectively. A subset of 64 primer pairs flanking SSR loci was used for screening polymorphism between two chickpea cultivars BG 256 and WR 315, which are parents of a *Fusarium* wilt mapping populations. Of them, 45 SSR markers (70.3%) were polymorphic between these two parents.

Key words: Microsatellites, ESTs, polymorphism, chickpea

Introduction

Chickpea (*Cicer arietinum* L.) has comparatively smaller genome size (750 Mbp) and shows only low levels of genetic polymorphism [1]. A large number of polymorphic markers are required for studying diversity or linkage analysis in this crop. Earlier studies in chickpea using RAPD and RFLP markers revealed limited polymorphism [2]. Simple Sequence Repeat (SSR) markers have been introduced in early nineties and are now being used widely in genetic characterization and diversity analysis in crop plants. It has been recently demonstrated that SSR markers can be useful tool for studying phylogenetic relationship and parental polymorphism within *Cicer* species [2, 3].

Huttel *et al.* [3] and Winter *et al.* [4] have developed and characterized 174 microsatellite markers for chickpea. Later, another 10 markers from chickpea were developed [5]. More recently, a set of 233 microsatellite markers have been generated for chickpea by Lichtenzveig [6] and 13 new markers by Chaudhary *et al.* [7]. Out of the total 430 SSR markers reported in chickpea, 118 SSRs have been positioned in the inter-specific map generated from the cross *C. arietinum* with *C. reticulatum* [8] whereas another of 51 SSRs have been positioned on the intra-specific chickpea linkage map [9]. The number of SSR markers available for chickpea genome mapping is still limited and only low marker density has been achieved in the maps reported till date. Moreover, the reported microsatellite sequences from chickpea have been isolated either by conventional genomic library screening procedure [3, 4, 5] or generated from BAG libraries. The presence of SSRs in the transcripts of genes suggests that they may have a role in gene expression or function [10]. Expansion and contraction of SSR repeats in genes of known function, therefore, can be tested for association with phenotypic variation or, more desirably, biological function [11]. Although in earlier studies, SSRs were reported more in non-coding region of eukaryotes [12], a larger number of tri-nucleotide repeats have been reported in the coding regions of higher genomes [13, 14]. The development of SSR markers requires a great deal of time, effort and investment in construction and screening of genomic libraries and sequencing of clones containing

SSR, primer development and their validation. However, a large number of EST and genomic DNA sequences available in the public databases provides an alternative method of microsatellite development. SSRs can be computationally mined from EST and genomic DNA databases [14]. The present study was undertaken with the objectives to develop and validate SSR markers from large EST and genomic sequence collections of chickpea available in public domain and to study the frequency and distribution of SSRs within two chickpea parental lines.

Materials and methods

Search for microsatellites and primer design

Chickpea genomic DNA and EST sequence data available in the NCBI GenBank (http://www.ncbi.nlm.nih.gov) were downloaded in Fasta format. EST sequences less than 200 bp were not included in the analysis. The identification of microsatellites was carried out using the online search tool 'WebTroll' (http:www.wsmartins.net/webtroll/troll). The retrieved sequences in Fasta format was analyzed for presence of repeat patterns for all possible di-, tri-, tetra-, and penta-nucleotide repeats. Microsatellites greater than 18 bp were only considered for tri-nucleotide repeats, indicating presence of at least six repeat units. Similarly for di-nucleotides and tetra nucleotides number of repeat units were 20 and 5, respectively with corresponding length of 40 bp and 20 bp.

Plant material

Two cultivars of chickpea, BG 256 (susceptible to Fusarium wilt caused by *Fusarium oxysporum* f. sp. *ciceri*) and WR 315 (resistant to Fusarium wilt) were used for polymorphism survey. The seeds were washed in sterile water and germinated on soaked paper towel in etiolated condition for extraction of genomic DNA from the seedlings.

Isolation of genomic DNA

DNA from young seedlings was extracted based upon the modified protocol of Guillemant and Laurence [15]. Samples were crushed to a fine paste with the mortar and pestle using extraction buffer [100 mM sodium acetate, pH 4.8; 50 mM Tris, pH 8.0; 50 mM EDTA, pH 8.0; 500 mM Nacl; 2% PVP (MW 10000); 1.4% SDS] and incubated for 30 min in a water bath prewarmed to 65°C. Ammonium acetate (10M, 0.6 volumes) was added into each tube and kept at 65°C for another 15 minutes followed by centrifugation at 10K rpm (Sorvall RC 5C rotor) for 10 minutes. Extracted supernatant was mixed with 0.6 volume of pre-chilled iso-propyl alcohol and kept at –20°C for 60 minutes. Precipitated DNA was pelleted by centrifugation, washed twice with 70% ethanol and dissolved in $T_{10}E_1$ buffer (10 mM Tris, 1.0 mM EDTA, pH 8.0). Dissolved DNA solution was extracted with phenol: chloroform: iso-amyl alcohol (25:24:1). RNA was removed by RNAse treatment (@ 4 µl/ml of supernatant from stock of 10mg/ml of RNAse) at 37°C for 1 hr. RNAse treated DNA was further extracted twice with chloroform: iso-amyl alcohol (24:1), re-precipitated in chilled ethanol and dissolved in TE buffer. Purified DNA was checked for its quality and quantity by agarose gel electrophoresis (0.8%) using uncut lambda (l) DNA (100 ng/µl) as control. For use in PCR, DNA was finally diluted to 25 ng/µl.

SSR analysis

PCR amplification was carried out in 0.2 ml thin-wall PCR tubes using the DNA Engine (MJ Research, model PTC 200). A total of 64 SSR primers were screened in the present study in which 63 SSR primers showed amplification. Polymerase chain reaction (PCR) mixture of 20 p.1 contained 25 ng of genomic DNA template, 0.6 U of *Taq* DNA polymerase (Bangalore Genei, Bangalore, India), 0.3 µM each of forward and reverse primers. 2.0 µl of 10 X PCR assay buffer (50 mM KCl, 10 mM Tris-Cl, 1.5 mM $MgCl_2$) and 0.25 µl of pooled dNTPs (100 mM each of dATP, dCTP, dGTP and dTTP from (MBI Fermentas, USA). PCR cycle conditions while performing SSR were as follows: initial denaturing step at 94°C for 3 min followed by 39 cycles of 94°C for 1 min, 50-55 C for 1 min and 72°C for 2 min. In the last cycle, primer extension at 72°C for 7 min was provided.

PCR products were electrophoretically separated on a 2% agarose gel, containing ethidium bromide using 1X TBE buffer (pH 8.0). The amplified products were visualized and photographed under UV light source. 100bp DNA Ladder (Bangalore Genei) was used as molecular weight marker for analysis of SSR. The product sizes were analyzed by AlphaEaseFC software version 3.1.2 (Alpha Innotech Corporation, USA). The PCR products from SSR analyses were scored quantitatively for presence or absence of amplicons. DNA bands were scored '1' for its presence and '0' for its absence for each primer-genotype combination. The scores were analyzed for estimating the polymorphism % and also to identify the relative efficiency of different repeats in revealing polymorphism.

Table 1. List of SSR markers showing polymorphism in chickpea

Name	Repeat	Primer	Length (bp)	NCBI GI No.	Size in BG256	Size in WR315
SSR 1	$(AG)_{20}$	TGAATTTTGTTTACCACCCCTC F TTTGGCTTATTCTGTTCTTCCC R	157	77745195	175	175
SSR 2	$(ATT)_6$	GACAAAACAACCTCCCAAGAAA F GACGACAACAACAACAACAACA R	258	110681821	232	244
SSR 4	$(TTG)_6$	GACAAAACAACCTCCCAAGAAA F AACAACGACAACAACAACAACG R	279	110681821	229	247
SSR 5	$(AAAT)_5$	GAGCCCTGAAATGAAGAAAGAA F CACCTTTGAGCCCTAGTCTGTT R	387	110681821	322, 295	306
SSR 7	$(CA)_{33}$	CACACACACAGACACACACACA F TGGTTCAGACATCACACCAAAT R	136	42405284	193, 111	189, 108
SSR 13	$(GA)_{36}$	ATACGACGACGATTCTGGATTT F TTCTCACATCTCTCTCTCTCTCTCTC R	170	42405133	181	174
SSR 14	$(GA)_{39}$	ACCTCCGTCCACACATTTCTAC F GTCGAAGCCATTGTTTTGTTG R	224	42405103	200	249
SSR 21	$(GA)_{20}$	GGGCCATACATCAAACACAAT F CCACATTCTTTAGCACATGGAA R	249	42404742	258	239
SSR 22	$(AATG)_5$	GCTTTCCCTTTACTTCTTGGGT TGCTATTCAAGTCTCCCTCCTC R	275	71153856	304	319
SSR 24	$(TAT)_7$	TGTCAGTGGATCACCAATTAGC F CAATCCCCATAAGATGAACTCC R	354	10241559	380	305
SSR 25	$(AG)_{20}$	GGCACATGGTTCCTCTTAAACT F CCATCACACCTTATTGCTTTCA R	352	33088334	405	411, 368
SSR 26	$(GA)_{23}$	TGAAGTCGTGGTGAAGAAAGAG F TGACTCCCGCATCATCATT R	214	33088322	329, 219	322, 202
SSR 28	$(CAG)_6$	AACTCGGGAGAGATGTTCTTCA F GGCTCTTTGCATGTTGCTG R	386	5262140	453	438
SSR 31	$(AAC)_{14}$	TAACGACAACGACAACAACAGC F GCCATTCCAGAGAGCCTTG R	161	5262137	379	382
SSR 35	$(CAA)_9$	TGCTCATGCTTACTTCTTCCTG F GTCGGCTTGGCTCATGTAAT R	220	5262132	348, 212	364, 87
SSR 38	$(TCT)_6$	GCGAAGGCTGTTAAGTGGAG F GACGACGAGGATGAGGACTT R	334	5262122	552	612
SSR 39	$(CCAACA)_7$	CCTGTGCATAAAGAAAACCTCC F GGTAGAAACGACGAATAGGGC R	149	5262121	822	865
SSR 42	$(GAT)_7$	CCTTTGGGTGGTTCATAGAAAA F CATCGTAATCATCGGTACTCCA R	229	7635493	268	268
SSR 46	$(TTG)_8$	GGTGTTGTTGTTGTTGTTGTTG F GCTTAAACCTACCCCTAATACGAA R	258	5262127	843	942
SSR 56	$(AG)_{20}$	GTGTGAGGAAAATTGAGGGAAG F ATGATTACGCCAAGCTCAGAAT R	251	47832810	748, 262	757, 262
SSR 58	$(CT)_{20}$	GATTCGCCCTTTCGAGCG F TGGTGAGAGAAGCAAGACCCAT R	156	47832732	123	280
SSR 60	$(AAAT)_6$	GGTCATGTTGATTTCTCACCAA F GAACTTTCCGCACACGTTATG R	337	87253106	425, 154	400, 142
SSR 61	$(ATA)_{14}$	GTTACAAGTCGCCATTCCAAA F CATTTGTCTCGTTCACATACCG R	398	33186957	684, 109	684, 109
SSR 62	$(AT)_{32}$	CATGCTCCCCTAATTGACATA F AGGCATAAATCCATCTGCATAC R	355	33186957	427, 270	438, 263
SSR 63	$(TTA)_7$	TCCGATGGAACCTTCTCTTTTA F CTCTTCGGGGTCGTATTGATT R	396	32490613	446	432

Results and discussion

The present study used "WebTroli' program to identify SSR sequences from 2100 ESTs and genomic DNA sequences that had been submitted to the NCBI GenBank database (http://www.ncbi.nlm.nih.gov/Genbank). Tandem Repeat Occurrence Locator (TROLL), is a light-weight SSR finder based on a slight modification of the Aho-Corasick algorithm [16]. It is a kind of dictionary-matching algorithm that locates elements of a finite set of strings (the "dictionary") within an input text. It is fast and requires a standard Personal Computer (PC) to operate. Various primer-parameters such as GC content, melting temperature (T_m), self annealing, end annealing, etc., had analyzed while designing primers. Many general primer designing computer tools can quickly perform such analysis to suggest the optimum primer pair within a specified region of nucleotide sequence. Oligonucleotides were designed to have a melting temperature between 55°C and 62°C to minimize formation of undesired hairpin or slipped-duplex conformations. The %GC content of the primers was kept between 40 and 60% and the product size was set at 100-400 base pair.

The program identified 406 SSR-containing sequences from the 2100 analyzed sequences'. Among the di-nucleotide motifs, GA/AG/CT was the most abundant. From the 406 SSR motifs identified, primers for 64 sequences could be designed. Except one primer pair, all SSR primer pairs amplified genomic DNA of two chickpea cultivars (BG 256 and WR 315). Out of 63 primers, 45 (70.31%) showed polymorphism and 18 yielded monomorphic bands. These two commercial chickpea genotypes were chosen for screening of the designed primers, because they are the parents of a population currently being used for linkage mapping and quantitative trait loci (QTL) tagging for the wilt resistance gene. Primer sequences, polymorphism, repeats and expected product size for the 25 highly polymorphic primers pairs are given in Table 1.

The proportion of polymorphic primers was highest for tetra-nucleotide repeats as all the three primers showed polymorphism. This was followed by tri-nucleotide repeats, where 20 out of 29 primers resulted in polymorphic amplifications. The polymorphism was least for penta-nucleotide repeats where only one out of three primers was polymorphic. In case of di- and hexa-nucleotide repeats, 19 and one primer were polymorphic. The size of amplicons ranged from 73 to 811 bp. It has been reported that EST-derived SSRs are less polymorphic than those derived from genomic libraries [14]. Some of the primers amplified larger fragments than the expected size, reflecting the possible presence of introns within the genomic DNA sequence.

Microsatellite markers are important for chickpea research because they are PCR based markers and easy to perform. They are the products of specific primers more stable than those generated by random primers such as Random Amplified Polymorphic DNA (RAPD) markers. Another advantage is their locus-specificity and transferability across genotypes within the species, which is an important feature for mapping purpose. Another possible advantage of using EST-derived SSR markers is that once mapped, they will always be associated with the genes carrying them. In fact, many SSR harboring ESTs show homology to the known genes when used for searches with BLAST-X (http://www.ncbi.nlm.nih.gov/BLAST/Blast.cg). SSR technology offers the potential of more cost effective data acquisition than other marker technologies. SSRs amplification profile varies in different varieties of chickpea, and this variability may be used to develop molecular markers for mapping important genes and traits in chickpea. These markers .being the derivative of gene sequences are expected to be of immediate use in molecular marker assisted breeding.

References

1. **Kazan K. and Muehlbauer F. J.** 1991. Allozyme variation and phylogeny in annual species of *Cicer* (Leguminosae). Plant Syst. Evol., **175**: 11-21.
2. **Udupa S. M., Robertson L. D., Weigand F., Baum M. and Hahl G.** 1999. Allelic variation at (TAA)n Microsatellite loci in a world collection of chickpea (*Cicer arietinum* L.) germplasm. Mol. Gen. Genet., **261**: 354-363.
3. **Huttel B., Winter P., Weising K., Choumane W., Weigand F. and Kahl G.** 1999. Sequence-tagged microsatellite site markers for chickpea (*Cicer arietinum* L). Genome, **42**: 210-117.
4. **Winter P., Pfaff T., Udupa S. M., Huttel B., Sharma P. C., Sahi S., Arreguin-Espinoza R., Weigand F., Muehlbauer F. J. and Kahl G.** 1999. Characterization and mapping of sequence-tagged microsatellite site in the chickpea (*Cicer arietinum* L.) genome. Mol. Gen. Genet., **262**: 90-101.
5. **Sethy N. K., Shokeen B. and Bhatia S.** 2003. Isolation and characterization of sequence tagged-microsatellite markers in chickpea (*Cicer arietinum* L.). Mol. Ecol. Notes, **3**: 418-430.
6. **Lichtenzveig J., Scheuring C., Dodge I., Abbo S. and Zhang H. B.** 2005. Construction of BAG and BIBAC libraries and their applications for generation

or SSR markers for genome analysis or chickpea, *Cicer arietinum* L. Theor. Appl. Genet., **110**: 492-510.

7. **Choudhary S., Sethy N. K., Shokeen B. and Bhatia S.** 2006. Development of sequence-tagged microsatellite site markers for chickpea (*Cicer arietinum* L.). Mol. Ecol. Notes, **6**: 93-95.
8. **Winter P., Benko-Iseppon A. M., Huttel B., Ratnaparkhe M., Tullu A., Sonnante G., Pfaff T., Tekeoglu M., Santara D., Sant V. J., Rajesh P. N., Kahl G. and Muehlbauer F. J.** 2000. A linkage map of chickpea (*Cicer arietinum* L.) genome based on recombinant inbred lines from a *C. arietinum* x *C. reticulatum* cross: localization of resistance gene, for *fusarium* wilt races 4 and 5. Theor. Appl. Genet, **101**: 1155-1163.
9. **Flardez-Galvez H., Ford R., Pang E. C. and Taylor P. W.** 2003. An intraspecific linkage map of the chickpea (*Cicer arietinum* L.) genome based on sequence tagged microsatellite site and resistance gene analog markers. Theor. Appl. Genet., **106**: 1447-1456.
10. **Cummings C. J. and Zoghbi H. Y.** 2000. Trinucleotide repeats: Mechanisms and pathophysiology. Ann. Rev. Genomics Hum. Genet., **1**: 281-328.
11. **Ayres N. M., McClung A. M., Larkin P. D., Bligh H. F. J., Jones C. A. and Park W. D.** 1997. Microsatellites and a single nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. Theor. Appl. Genet., **94**: 773-781.
12. **Hancock J. M.** 1995. The contribution of slippage-like processes to genome evolution. J. Mol. Evol., **41:** 1038-1047.
13. **Li Y. C., Korol A. B., Fahima T. and Nevo E.** 2004. Microsatellites within genes: structure, function and evolution. Mol. Biol. Evol., **21:** 991-1007.
14. **Varshney R. K., Graner A. and Sorrells M. E.** 2005. Genie microsatellite markers in plants: features and applications. Trends Biotechnol., **23**: 48-55.
15. **Guillemant P. and Laurence M. D.** 1992. Isolation of Plant DNA: A fast, inexpensive, and reliable method. Plant Mol. Biol. Rep., **10**: 60-65.
16. **Aho Alfred V. and Margaret J. C.** 1975. Efficient string matching: An aid to bibliographic search. Communications of the ACM, **18**: 333-340.